Incorporating Discourse Information within Textual Entailment Inference

Shachar Mirkin

November 2010

Charles University, Prague

Joint work with:

Jonathan Berant , Ido Dagan, Sebastian Padó, Eyal Shnarch

Outline

- Introduction
  - Discourse and Inference
  - Textual Entailment
    - Entailment recognition process
    - The RTE-5 Search task dataset
- The role of discourse references in entailment inference (ACL-10)
  - Theoretical analysis
- Recognising entailment within discourse (COLING-10)
  - Empirical study
- Conclusions

Introduction

Discourse and Inference

The Russian navy worked desperately on Friday to save a small military submarine. The seven men on board were said to have as little as 24 hours of air.

- The 1st sentence is needed for the interpretation of the 2nd
  - *small military submarine* clarifies who and where the *seven men* are
- *How many people were trapped on board the submarine?*
  - Discourse-sensitive inference is needed to integrate information across sentences
- Useful for inference-based NLP applications
  - QA, IE, IR, text comprehension, etc.

Textual Entailment (TE):

- A generic framework for applied semantic inference
- Can the meaning of a target textual assertion (*hypothesis*, H) be inferred from a given *text* (T)?

H *The mini-submarine was trapped underwater*

T *It became tangled up with a metal cable or a sunken trawler*

- In this case: ***T entails H ($T \Rightarrow H$)***
- TE makes an appropriate framework for investigating inference-related phenomena

Discourse is relevant again

Typical entailment recognition flow:

1. **Preprocessing**
   - …
   - Coreference resolution

2. **Matching T & H components *to obtain maximal coverage***

   Using knowledge-based entailment resources:
   - WordNet, DIRT, CBC, Wikipedia, …

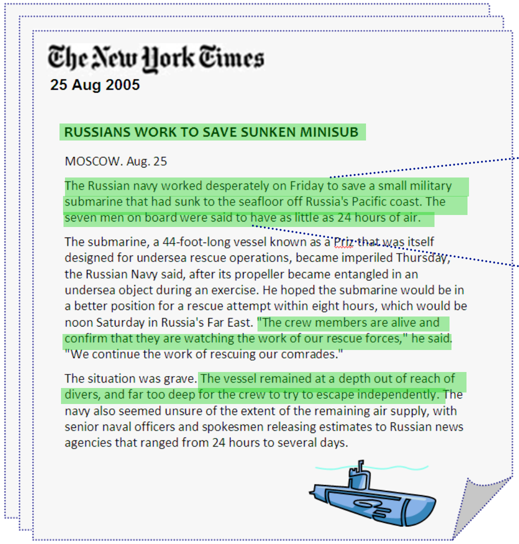
   I. Transformations: Modify T making it more similar to H
      - $T \rightarrow T_1 \rightarrow T_2 \ldots \rightarrow T_n$ (s.t. $T_i \Rightarrow T_{i+1}$)

   II. Alignment between T and H components

3. **Approximate matching**
   - Bridge knowledge gaps
     - Cost functions, classification, etc.

A Discourse-sensitive Entailment Setting

- **RTE-5 "Search" task**
  - Find all sentences in the corpus that entail H
  - Each sentence (T) is judged individually (does T entail H?)
    - But is interpreted based on the entire discourse



The New York Times
25 Aug 2005

RUSSIANS WORK TO SAVE SUNKEN MINISUB

MOSCOW. Aug. 25

The Russian navy worked desperately on Friday to save a small military submarine that had sunk to the seafloor off Russia's Pacific coast. The seven men on board were said to have as little as 24 hours of air.

The submarine, a 44-foot-long vessel known as a Priz that was itself designed for undersea rescue operations, became imperiled Thursday, the Russian Navy said, after its propeller became entangled in an undersea object during an exercise. He hoped the submarine would be in a better position for a rescue attempt within eight hours, which would be noon Saturday in Russia's Far East. "The crew members are alive and confirm that they are watching the work of our rescue forces," he said. "We continue the work of rescuing our comrades."

The situation was grave. The vessel remained at a depth out of reach of divers, and far too deep for the crew to try to escape independently. The navy also seemed unsure of the extent of the remaining air supply, with senior naval officers and spokesmen releasing estimates to Russian news agencies that ranged from 24 hours to several days.

*on board **the sunken submarine***

T: *The seven men on board were said to have as little as 24 hours of air.*

$\Downarrow$?

H: *The mini submarine was trapped underwater*

- Discourse is very relevant
  - But almost completely ignored by the task participants

Assessing the role of discourse references in entailment inference

(Mirkin, Dagan, Padó, ACL-2010)

Discourse References

We consider two types of *discourse references*:

- **Coreference**
  - Indicating reference to the **same** entity or event
  - *I entered the **room**. **It** was bright and cozy.*

- **Bridging** **(Clark, 1975)**
  - An **implied relation** between two **distinct** entities or events
  - *I've just **arrived**. The **camel** is outside and needs water*

    (Asher and Lascarides, 1998)
  - *I entered the **room**. The **ceiling** was very high.*

Motivation & Goals

- Discourse useful in NLP inference (as captured by TE)
- So far: limited integration of discourse into entailment
  - Substitution of nominal coreferents at preprocessing time
  - Showed little improvement

Is this the best that can be done?

- Our goals:
  - Semantic inference researchers:
    - How should inference systems utilize discourse information?
      - Is substitution enough?
  - Developers of reference resolvers:
    - What other types of discourse references are useful for inference?
      - Verbal references? Bridging?
    - Does current performance suffice?

Analysis Methodology

- Analyzed a sample of T-H pairs from the Search task:
  - Representation: dependency parse trees
  - Only references *relevant* for entailment: those that increase *coverage*

Target component

H: *Massachusetts allows homosexual marriages*

Reference term

T': *Spain will join the Netherlands and Belgium in allowing homosexual marriages.*

…

Focus term

T: *Such unions are also legal in the northeastern US state of Massachusetts.*

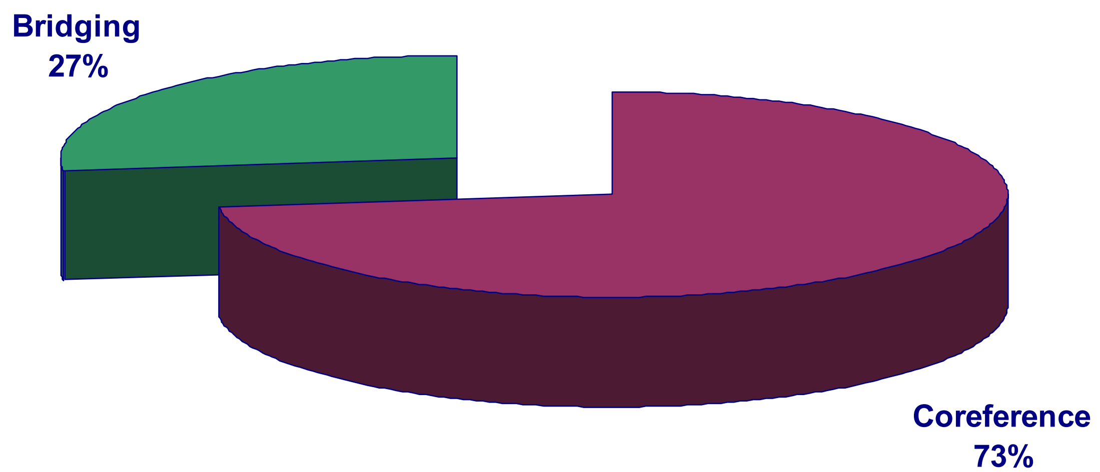
- Annotated reference elements:
  1. ***Target component*** in H – a component not directly covered by T
  2. ***Focus term*** in T
  3. ***Reference term*** - in T's discourse
     - In a reference relation with the focus term
     - Can help cover the target component
- By identifying the reference, the target component can be inferred

General statistics

- Analyzed 120 sentence-hypothesis pairs
- Relevant references found in 72% of the pairs
  - Many with multiple ones
  - In total: 137 relevant references

➢ Discourse references do play an important role in inference

Reference Types



- Bridging relations:
  - Cause: e.g. between *injured* & *the attack*
  - Location of events, event participants , set membership, etc.

➢ Coreference is prominent but bridging should also be addressed

Syntactic Types

(%)	Pronoun	NE	NP	VP
Focus term	9	19	49	23
Reference term	-	43	43	14

- Nominal phrases are dominant

- Almost a quarter of the focus terms are verbal phrases
  - E.g. *(the ice is) softening – melting*
  - Frequently crucial for entailment -- include the main predicate of H

➢ Verbal references can't be overlooked

Integrating Discourse References into Inference

- How to integrate into the inference process?

- Different ways to cover the target component

- 3 cases:

  1. A textual fragment covering the **entire** target component is found elsewhere in the text
     - **Substitution**
  2. Needed fragments are scattered among **multiple locations** in the text
     - **Merge**
  3. The target component is only **implicitly** implied in the text
  - e.g. via a **bridging** relation
     - **Insertion**

- 3 operations over the focus & reference to achieve coverage

Substitution

Target component

H: *Massachusetts allows homosexual marriages*

T' : *Once the reform becomes law, Spain will join the Netherlands and Belgium in allowing homosexual marriages*

Reference term

…

Focus term

T: *Such unions are also legal in six Canadian provinces and the northeastern US state of Massachusetts.*

Substitution: replace the focus term in T with the reference term from T' :

$T \rightarrow T_1$: ***Homosexual marriages*** *are also legal in six Canadian provinces and the northeastern US state of Massachusetts.*

- Representing state of the art usage of discourse in inference systems

Merge

Target component

H: *The* ***Russian mini submarine*** *was trapped underwater*

T': *All seven aboard the Russian submarine appeared to be in satisfactory condition, naval spokesman said.*

Reference term

...

T: *The military was racing against time early Friday to rescue the mini submarine trapped on the seabed.*

Focus term

Substitution?

T→T_1: *The military was racing against time early Friday to rescue the* ***Russian submarine****.*

- … but now ***mini*** isn't covered

Merge: merge the modifiers of the focus & reference terms under a single head

T→T_1: *The military was racing against time early Friday to rescue the* ***Russian mini submarine*** *trapped on the seabed.*

Insertion

Target component

H: *A recent accident **in China** has killed several miners.*

T' :*China seeks solutions to its coal mine safety.*

Reference term

…

Focus term

T: *A recent accident has cost more than a dozen miners their lives.*

Insertion: Insert into T the implied relation in order to connect the reference term to the focus term

$T \rightarrow T_1$: *A recent accident in **China** has cost more than a dozen miners their lives.*

Inserted relation

- A specific insertion for each bridging relation
  - Bridging type needed as output of the bridging resolver

Operations

(%)	Sub	Merge	Insertion
Coreference	62	38	-
Bridging	30	-	70
Total	54	28	18

- **Substitution** is most frequent
  - But in nearly half the cases, a different operation is needed
- **Merge** – complete missing modifiers in both NPs and VPs
- **Insertion** – majority of bridging relations

➢ Incorporating other discourse operations in inference

Discourse-based knowledge vs. entailment knowledge (1)

H: *The ice is **melting** in the Arctic*

T: *The scene at the **receding** edge of the Exit Glacier was part festive gathering, part nature tour with an apocalyptic edge.*

To cover ***melting***:

- Option 1: use inference rule: ***receding ↔ melting***
  - context-dependent

- Option 2: use discourse to detect event coreference

T' : *. . . people moved closer to the rope line near the glacier as it shied away, practically groaning and **melting** before their eyes.*

Discourse-based knowledge vs. entailment knowledge (2)

H: *The BTK **serial killer** was accused of at least 7 killings starting in the 1970's.*

T: *Police say **BTK** may have killed as many as 10 people between 1974 and 1991.*

To cover *serial killer*:

- Option 1: use world knowledge
  - A person who killed 10 people over a period of time is a serial killer
- Option 2: Identify coreference between:
  - ***BTK*** in T & ***BTK serial killer*** in the discourse around T

T': *Rader, whom police believe is the **BTK serial killer**, wasn't well liked by his neighbors*

Implications:

- Discourse references can often provide world knowledge on the fly
  - Sometimes it's easier to resolve the reference
- Reference resolution may be used:
  - To learn new knowledge (entailment rules)
  - Complementarily to rules – to support existing ones

Conclusions – Discourse References Roles in Inference

- Our goals:
  - How should *inference* methods utilize discourse references?
  - How should *discourse* methods be extended to support inference?

- Current methods do not suffice:
  - Nominal coreference is prominent but:
    - Verbal references & bridging relations are abundant
  - Substitution alone is insufficient
    - Merge, insertion

- Discourse references and entailment knowledge can be used complimentarily

Recognising entailment within discourse

(Mirkin, Berant, Dagan, Shnarch, COLING-2010)

Motivation & Goals

- Texts are commonly interpreted based on the entire discourse
- Most entailment systems have addressed discourse only by:
  - Applying nominal term substitution
    - Via off-the-shelf coreference resolvers
- **Assumption:**
  - Discourse can play a much more important role
- **Goals:**
  - Identify additional types of discourse information relevant for inference
  - Use them to improve entailment performance

BIUTEE: *The Baseline RTE System*

BIUTEE (**B**ar-**I**lan **U**niversity **T**extual **E**ntailment **E**ngine) (Bar Haim et al., 2008)

- Basis for discourse-based enhancements & baseline

BIUTEE' s entailment recognition flow:

1. **Preprocessing**
   - MINIPAR dependency parse trees, stored in a *CompactForest*
   - OpenNLP coreference resolver

2. **Matching T & H components**
   - Transformation-based approach
   - Knowledge-based entailment transformations
     - WordNet , DIRT & Syntactic (e.g. passive to active)

3. **Approximate matching**
   - Supervised classification
   - Features to measure the coverage of H components by T
     - Nouns, predicates, named entities, dependency edges etc.

Addressing discourse

- Coreference
- Bridging
- Coherence

Based on our discourse references analysis: significant for inference

Coreference: *Augmented Coreference Set*

Observations:

- Many coreference relations include lexical match (well known)
  - E.g. *the Airbus A380's* ***first flight*** & ***first*** *test* ***flight***
- Coverage of available coreference tools is limited
  - Missing even on these 'simple' cases

Solution:

- Consider each pair of NPs in a document as coreferring if:
  1. Their heads are identical
  2. No semantic incompatibility is found between their modifiers

- Semantic incompatibility of modifiers
  i. Antonyms
  ii. Different numbers
  iii. Co-hyponyms:
    - *Greek* (gods) vs. *Roman* (gods)
    - *walking (distance)* vs. *running (distance)*

Bridging: *Global Information*

Observation:

- Prominent information is often assumed "globally" known throughout the text
  - Typically corresponds to ***bridging*** relations

Title:	***CHINA****'S COAL MINE DEATHS UP 8.5 PERCENT*
3:	*Coal mine fatalities increased 8.5 percent in 2005.*

  - *China* isn't mentioned in (3), but is assumed known (implicitly implied)

- Tools for bridging resolution are unavailable

Method:

- Identify key terms in the document
  - Add them **explicitly** to each sentence
  - Key terms identification:
    - tf-idf with extra weight to named entities (esp. locations) & title terms

Coherence: *Considering the Previous Sentence*

Observation:

- Adjacent sentences refer to the same entities & events


The New York Times
25 Aug 2005

RUSSIANS WORK TO SAVE SUNKEN MINISUB

MOSCOW. Aug. 25

The Russian navy worked desperately on Friday to save a small military submarine that had sunk to the seafloor off Russia's Pacific coast. The seven men on board were said to have as little as 24 hours of air.

The submarine, a 44-foot-long vessel known as a Priz that was itself designed for undersea rescue operations, became imperiled Thursday, the Russian Navy said, after its propeller became entangled in an undersea object during an exercise. He hoped the submarine would be in a better position for a rescue attempt within eight hours, which would be noon Saturday in Russia's Far East. "The crew members are alive and confirm that they are watching the work of our rescue forces," he said. "We continue the work of rescuing our comrades."

The situation was grave. The vessel remained at a depth out of reach of divers, and far too deep for the crew to try to escape independently. The navy also seemed unsure of the extent of the remaining air supply, with senior naval officers and spokesmen releasing estimates to Russian news agencies that ranged from 24 hours to several days.

1: *The Russian navy worked desperately on Friday to save a small military submarine.*

2(+1): *The Russian navy worked desperately on Friday to save a small military* ***submarine****. The*

2: *The seven men on board were said to have as little as 24 hours of air.*

H: *The mini submarine was trapped underwater*

Method:

- Consider (1) when classifying (2)
  - to improve the coverage of H by (2)
  - *submarine* is covered
- Generate features from current & previous sentences together
  - in addition to 'regular' features

Coherence: *Entailment Bulks*

Observation:

- Entailing sentences tend to be adjacent to each other

The New York Times
25 Aug 2005

RUSSIANS WORK TO SAVE SUNKEN MINISUB

MOSCOW. Aug. 25

The Russian navy worked desperately on Friday to save a small military submarine that had sunk to the seafloor off Russia's Pacific coast. The seven men on board were said to have as little as 24 hours of air.

The submarine, a 44-foot-long vessel known as a Priz that was itself designed for undersea rescue operations, became imperiled Thursday, the Russian Navy said, after its propeller became entangled in an undersea object during an exercise. He hoped the submarine would be in a better position for a rescue attempt within eight hours, which would be noon Saturday in Russia's Far East. "The crew members are alive and confirm that they are watching the work of our rescue forces," he said. "We continue the work of rescuing our comrades."

The situation was grave. The vessel remained at a depth out of reach of divers, and far too deep for the crew to try to escape independently. The navy also seemed unsure of the extent of the remaining air supply, with senior naval officers and spokesmen releasing estimates to Russian news agencies that ranged from 24 hours to several days.

Consequence:

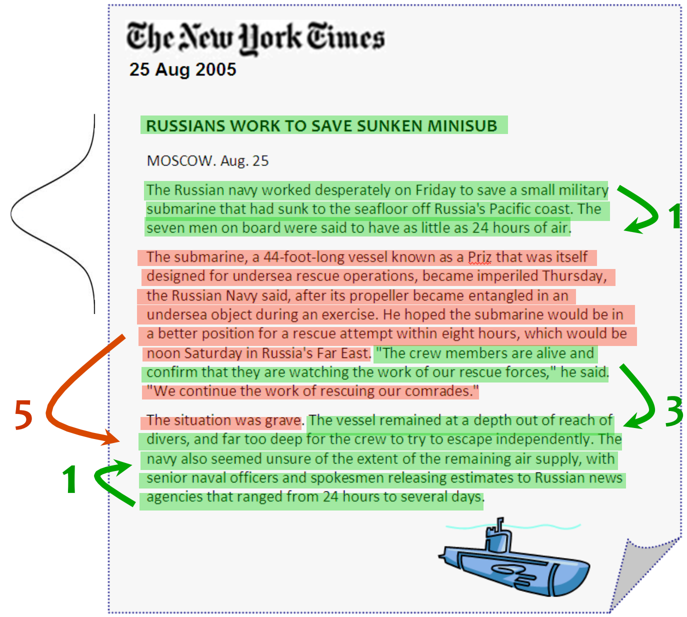
- If we identify an entailing sentence
  - it's more likely that its neighboring sentences are also entailing

Method:

- 2-phase classification scheme
  I. Classify sentences independently
  II. Consider classification of phase-I
    - Extract *meta-features*
    - Make final classification decision
- Allows using an ensemble of classifiers ("*base-classifiers*")

Meta-features: *Capturing Entailment Bulks*

- **Second-closest entailment**
  - The distance to the 2nd closet entailing sentence (inc. self)
  - Entailing sentences in bulks get a lower distance



- **Smoothed entailment**
  - Smoothing the classification score with scores of adjacent sentences
    - weighted by their distance
  - Boosting scores of sentences in entailment environment

Meta-features: *Title Entailment*

- **Title entailment**
  - If the title is entailing, the document is more likely to contain entailing sentences

The New York Times
25 Aug 2005

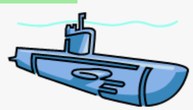
RUSSIANS WORK TO SAVE SUNKEN MINISUB

MOSCOW. Aug. 25

The Russian navy worked desperately on Friday to save a small military submarine that had sunk to the seafloor off Russia's Pacific coast. The seven men on board were said to have as little as 24 hours of air.

The submarine, a 44-foot-long vessel known as a Priz that was itself designed for undersea rescue operations, became imperiled Thursday, the Russian Navy said, after its propeller became entangled in an undersea object during an exercise. He hoped the submarine would be in a better position for a rescue attempt within eight hours, which would be noon Saturday in Russia's Far East. "The crew members are alive and confirm that they are watching the work of our rescue forces," he said. "We continue the work of rescuing our comrades."

The situation was grave. The vessel remained at a depth out of reach of divers, and far too deep for the crew to try to escape independently. The navy also seemed unsure of the extent of the remaining air supply, with senior naval officers and spokesmen releasing estimates to Russian news agencies that ranged from 24 hours to several days.



Title: *RUSSIANS WORK TO SAVE SUNKEN MINISUB*

H: *The mini-submarine was trapped underwater*

1^{st} sentence entailing title

- The first sentence in a news article often entails the article's title

(Bensley & Hickl., 2008)

  - ➢ If the title is entailing, we assume the 1^{st} sentence is also entailing

The New York Times
25 Aug 2005

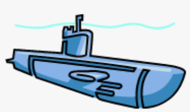
RUSSIANS WORK TO SAVE SUNKEN MINISUB

MOSCOW. Aug. 25

The Russian navy worked desperately on Friday to save a small military submarine that had sunk to the seafloor off Russia's Pacific coast. The seven men on board were said to have as little as 24 hours of air.

The submarine, a 44-foot-long vessel known as a Priz that was itself designed for undersea rescue operations, became imperiled Thursday, the Russian Navy said, after its propeller became entangled in an undersea object during an exercise. He hoped the submarine would be in a better position for a rescue attempt within eight hours, which would be noon Saturday in Russia's Far East. "The crew members are alive and confirm that they are watching the work of our rescue forces," he said. "We continue the work of rescuing our comrades."

The situation was grave. The vessel remained at a depth out of reach of divers, and far too deep for the crew to try to escape independently. The navy also seemed unsure of the extent of the remaining air supply, with senior naval officers and spokesmen releasing estimates to Russian news agencies that ranged from 24 hours to several days.



H: *The mini-submarine was trapped underwater.*

⇑

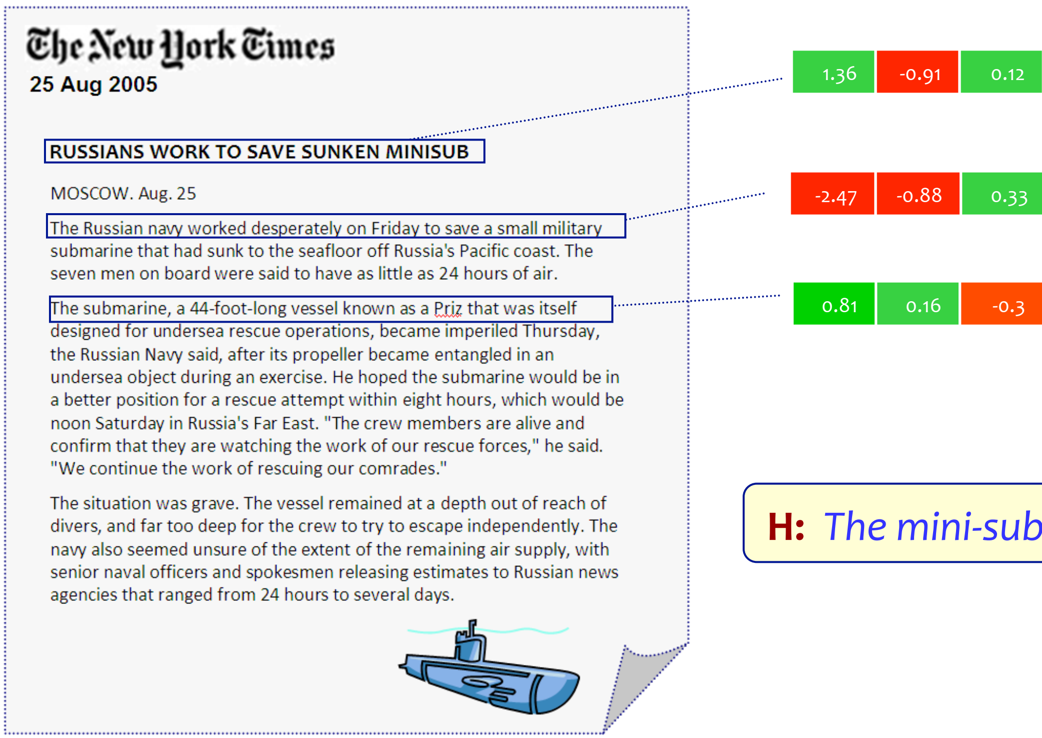
Title: *RUSSIANS WORK TO SAVE SUNKEN MINISUB*

⇑

1: *The Russian navy worked desperately on Friday to save a small military submarine that had sunk ..*

Meta-features: *Classification Scores*

- **Classification scores**
  - Of each of the base-classifiers


The New York Times
25 Aug 2005

RUSSIANS WORK TO SAVE SUNKEN MINISUB

MOSCOW. Aug. 25

The Russian navy worked desperately on Friday to save a small military submarine that had sunk to the seafloor off Russia's Pacific coast. The seven men on board were said to have as little as 24 hours of air.

The submarine, a 44-foot-long vessel known as a Priz that was itself designed for undersea rescue operations, became imperiled Thursday, the Russian Navy said, after its propeller became entangled in an undersea object during an exercise. He hoped the submarine would be in a better position for a rescue attempt within eight hours, which would be noon Saturday in Russia's Far East. "The crew members are alive and confirm that they are watching the work of our rescue forces," he said. "We continue the work of rescuing our comrades."

The situation was grave. The vessel remained at a depth out of reach of divers, and far too deep for the crew to try to escape independently. The navy also seemed unsure of the extent of the remaining air supply, with senior naval officers and spokesmen releasing estimates to Russian news agencies that ranged from 24 hours to several days.

H: *The mini-submarine was trapped underwater.*

Locality Meta-features: *Known Issue*

- Locality meta-features:
  - 1st sentence entailing title
  - 2nd closest entailment
  - Smoothed entailment

- *Known Issue:*
  - These features depend on the accuracy of base-classifiers

Results & analysis

Results (Micro Averaged)

- BIUTEE-DISC:
  - 3 base-classifiers : SVM^{perf}, Naïve Bayes, Linear regression

	P(%)	R (%)	F_1 (%)
All-yes baseline	4.6	100	8.9
BIUTEE-DISC	20.82	57.25	**30.53**
BIUTEE	14.53	55.25	23

- A low proportion of positives in the data

Results: *Is it only the ensemble?*

- BIUTEE-DISC & BIUTEE^{ensemble}:
  - 3 base-classifiers : SVM^{perf}, Naïve Bayes, Linear regression

	P(%)	R (%)	F_1 (%)
All-yes baseline	4.6	100	8.9
BIUTEE-DISC	20.82	57.25	**30.53**
BIUTEE	14.53	55.25	23
BIUTEE^{ensemble}	14.86	59	23.74

Results: *Removing locality features*

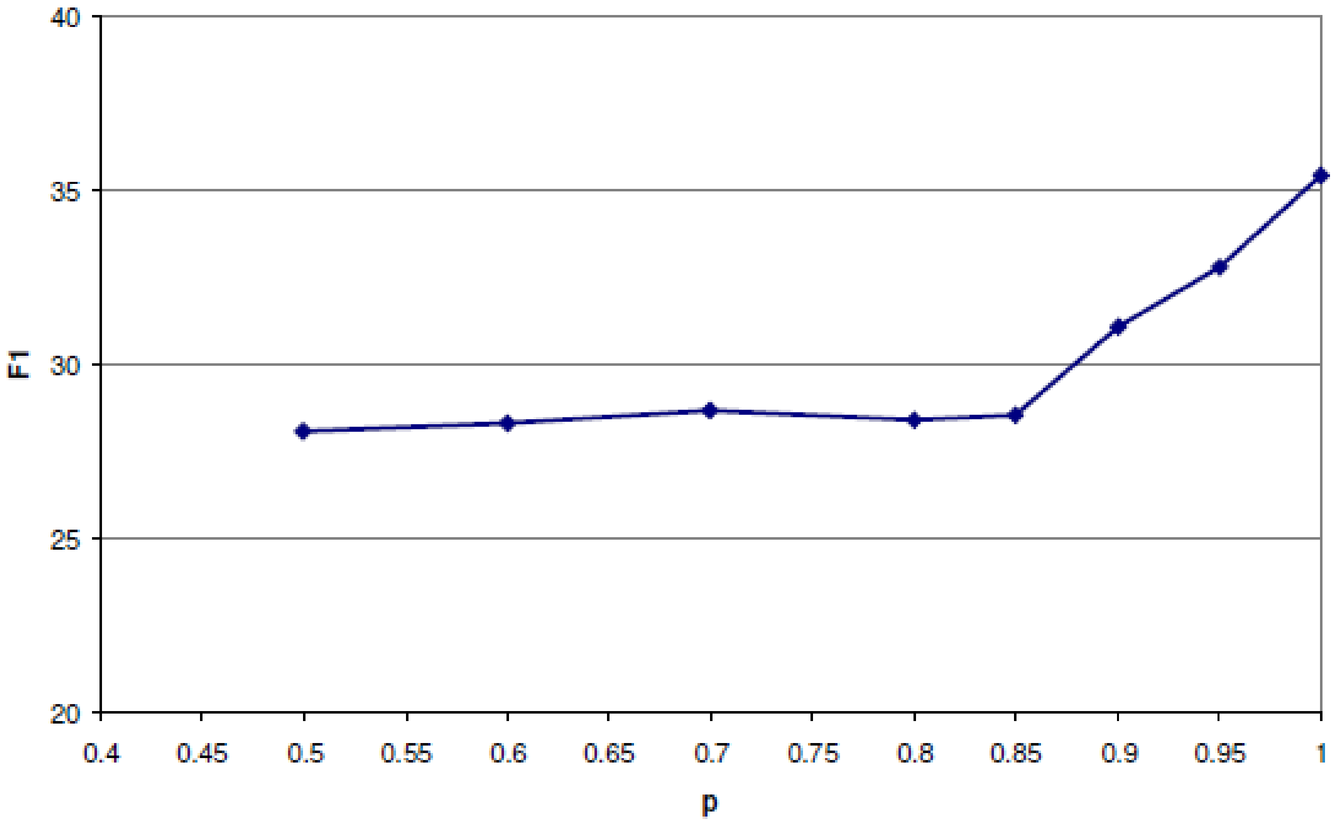
- BIUTEE-DISC$_{\text{no-loc}}$
  - No locality features

- Ensemble
  - Classification scores
  - Title entailment
- Previous sentence features
- Augmented coreference
- Global information

	P(%)	R (%)	F_1 (%)
All-yes baseline	4.6	100	8.9
BIUTEE-DISC	20.82	57.25	30.53
BIUTEE	14.53	55.25	23
BIUTEE$^{\text{ensemble}}$	14.86	59	23.74
BIUTEE-DISC$_{\text{no-loc}}$	22.35	57.12	**32.13**

Analysis of Locality Features

- Simulation, using an ***oracle classifier***
  - For locality features: the oracle provides the correct classification of **adjacent sentences** with probability *p*
  - Measuring F1 of BIUTEE-DISC in a range of values of *p*
    - Averaging 5 runs for each value of *p*



Ablation Tests

- Ablation tests relative to BIUTEE-DISC$_{\text{no-loc}}$ (F_1=32.13%)

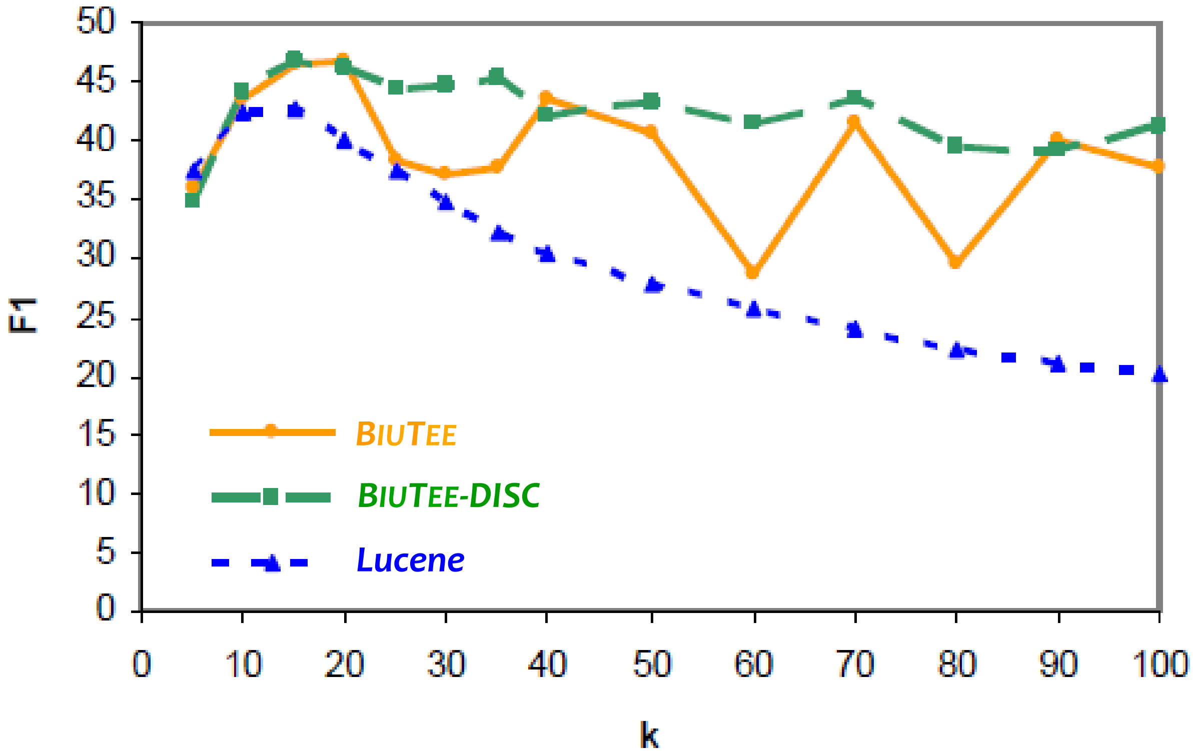
Removed Component	F_1 (%)	ΔF_1 (%)
Previous sentence features	28.55	3.58
Augmented coreference	26.73	5.4
Global information	31.76	0.37

Adding IR filtering

IR Filtering: *Motivation*

- Only 4% of development set T-H pairs are positive
- A candidate retrieval stage is called upon
  - Efficiency
  - RTE-5 systems applying IR-based filtering achieved best results

- BIUTEE-DISC vs. BIUTEE in this setting
  - In a range of top-*k* retrieved sentences



- BIUTEE-DISC vs. BIUTEE:
  - AUC is larger
  - More robust

Performance of Best Configurations

- Candidate retrieval setting
  - A single point of *k*, based on development set

		P(%)	R (%)	F_1 (%)
K=20	BIUTEE-DISC$_{\text{no-loc}}$	50.77	45.12	**47.78**
	BIUTEE$^{\text{ensemble}}$	51.68	40.38	45.33
K=15	Lucene	35.93	52.5	42.66
	RTE-5 best*	40.98	51.38	45.59
	RTE-5 second-best**	42.94	38	40.32

- Statistically significant (McNemar, $p<0.01$)

* Mirkin et al.,TAC 2009 (an earlier version of this work)

** MacKinlay and Baldwin, TAC 2009

Conclusions

Conclusions – Recognising entailment within Discourse

- First empirical investigation of discourse impact on entailment
  - Identified discourse phenomena relevant for inference
  - Suggested directions for incorporating them in the entailment process
    - Assessed their actual impact on performance
- Discourse provides useful information
  - Significantly improves entailment inference
  - Outperforming best systems on the task

Overall Conclusions and Future Work

Discourse & Inference: Conclusions and Future Work

- Conclusions

1. Better tools for discourse processing are required
2. Inference systems should incorporate discourse information in additional varied ways

- Future Work
  - Identify additional relevant discourse phenomena
  - Develop improved ways for utilizing discourse information in inference

Thank you!

Questions ?

Time for MT?

Source-Language Entailment Modeling for Translating Unknown Terms

(Mirkin et al., ACL 2009)

Motivation – Unknown Terms

- MT systems frequently encounter terms they are unable to translate - *unknown terms* (OOV)

- Particularly common for:
  - Languages-pairs for which parallel corpora are scarce
  - Different training-test domains

⇨ poor translation

Goal: improve translation of texts with unknown terms

➢ through entailment-based approach

Handling Unknown Terms – Baseline Approaches

- Translating to French:

  *"Cisco **filed a lawsuit** against Apple for patent violation"*

unknowns

Baseline approaches:

- Leaving the unknown terms **untranslated**

  *"Cisco **lawsuit filed** une contre Apple pour violation de brevet"*

- **Omitting** the unknown terms

  "Un Cisco contre Apple pour violation de brevet"

  (*"A Cisco against Apple for…"*)

Handling Unknown Terms – Paraphrasing

- Translating to French:

  *"Cisco **filed a lawsuit** against Apple for patent violation"*

unknowns

Paraphrasing (Callison-Burch et al., 2006)

- Translating a known paraphrase instead of the original term
- E.g.: *file a lawsuit* ⇔ *sue*

  Implicitly translating: *Cisco **sued** Apple for patent violation*

- Callison-Burch et al.'s implementation:
  - Requires multilingual corpora; noisy
  - Ambiguity is handled by the SMT-standard target LM

TE for MT

- **When paraphrases not available, generate source entailments**
- E.g.: *file a lawsuit* $\Rightarrow$ *accuse*

  *Cisco **filed a lawsuit** against Apple for patent violation* →

  *Cisco **accused** Apple for patent violation*

  - Improves coverage, still producing useful translations

- **Verify rule application with context models**

- **Use monolingual source-language Information:**
  - Monolingual resources & methods are more abundant
  - Better suited for directional rules

Textual Entailment for MT – Input & Output

Input

- A text *t* with one or more unknown terms
- A monolingual resource of entailment rules
- Pruning parameter *k*

Textual Entailment for MT

Output

A translation of either (in order of preference)

1. a paraphrase of *t*

   OR
2. a text entailed by *t*

   OR
3. *t* with unknown terms left untranslated

Textual Entailment for SMT – Method (Brief)

unknown

Cisco sued Apple

Not in phrase table

Source context-model scores

Target-model scores

Source-Target scores

~~sued ⇔ litigated~~ | -
sued ⇔ processed | 0.3
sued ⇒ accused | 0.6
sued ⇒ challenged | 0.1
… | …

Cisco processed Apple
Cisco accused Apple

Cisco a mixé une pomme
Cisco a accusé Apple

0.5
0.4

0.15
0.24

Pruning: k=2

- **SMT system:** Matrax (Simard et al., 2005)
- **Corpora** (from the shared translation task in WMT-2008):
  - **Training:** **Europarl** 1M English-French sentences
  - **Test:** **~2,500** **News** English sentences with unknown terms

- **Entailment rules resource:** WordNet 3.0
  - Rules for: nouns, verbs, adjectives, adverbs
  - **Paraphrases:** Synonyms (e.g. *provoke* ⇔ *evoke*)
  - **TE: adding directional entailments:** Hypernyms (*provoke* ⇒ *cause*)

Paraph ⊆ TE

- **Evaluation:**
  - Manual: Human annotators marking each translation for each model as acceptable or not
  - Automatic: BLEU, Meteor
    - although not suitable for semantic modifications

Model			Precision (%)		Coverage (%)	
	Src	Tgt	PARAPH.	TE	PARAPH.	TE
1	–	*SMT*				
2	*NB*	*SMT*				
3	*LSA*	*SMT*				
4	*NB*	–				
5	*FREQ*	–				
6	*RAND*	–				

Target-only model

Source-Target models

Source-only models

Baselines

- TE vs. Paraphrases: substantial coverage increase
  - with just a little decrease in precision
- *Src-Tgt* models (2-3) comparable to tgt-only (1), but more efficient
- Top models outperform the baselines

Comparison to Previous Approach

- Comparison to: Callison-Burch et al., 2006 **(CB)**
  - Phrase table augmentation using Europarl **parallel corpora**
- Manual evaluation (150 sentences): **acceptance** and **preference**

Model	Precision (%)	Coverage (%)	Better (%)
NB-SMT (TE)	85.3	56.2	72.7
CB	85.3	24.2	12.7

Conclusions

- A new approach for handling unknown terms in MT
- The first application of TE to MT for improving translation quality
- Translation can be improved using:
  - **Monolingual** resources
  - **Textual entailment** relationships
- Source context models allow performing the process efficiently
  - Rank candidates to allow pruning
- Further room for improvement
  - Developing better source & target context models